Анализ данных (Байесовские методы машинного обучения)

М. В. Лебедев

Московский авиационный институт

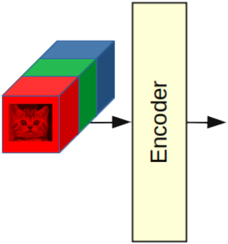
8 мая 2022 г.

Материалы к курсу

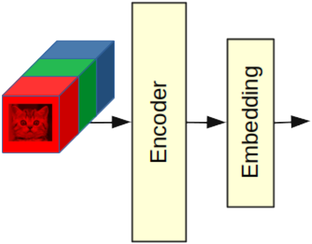
- Bayesian Methods for Machine Learning (Coursera)
- Байесовские методы машинного обучения
- Лекции Ветрова Д.П.
- Christopher M. Bishop. Pattern Recognition and Machine Learning

Сжатие размерности

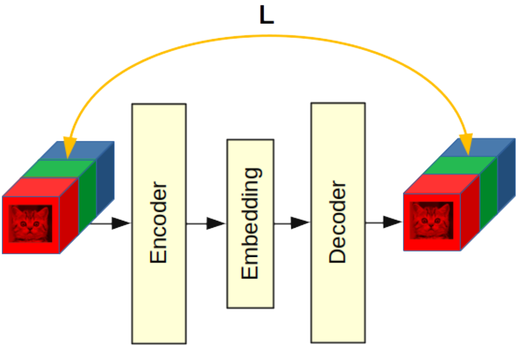
Сжатие размерности



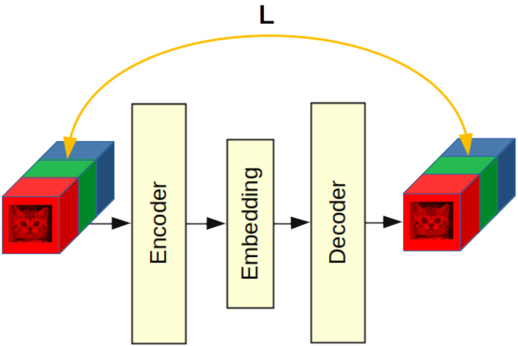
Сжатие размерности



Сжатие размерности

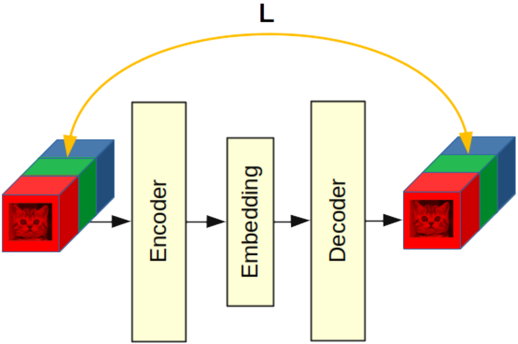


Сжатие размерности



$$L = \sum_{i=1}^{N} MSE(I_i, \hat{I}_i)$$

Сжатие размерности



$$L = \sum_{i=1}^{N} MSE(I_i, \hat{I}_i)$$

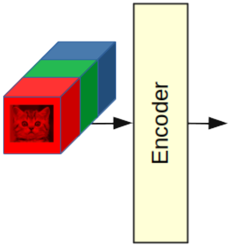
или

$$L = \sum_{i=1}^{N} BCE(I_i, \sigma(z_i)).$$

Сжатие размерности

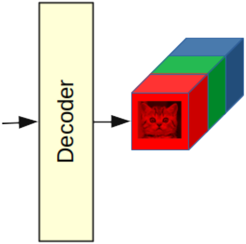
После обучения имеем два преобразования:

- *Encoder* : $\mathbb{R}^n \to \mathbb{R}^m, \quad (n \gg m)$.



Сжатие размерности

- $Decoder : \mathbb{R}^m \to \mathbb{R}^n, (n \gg m)$



При этом может происходить потеря качества:

$$Decoder(Encoder(x)) \neq x.$$

Масштабируемость вариационного вывода

Масштабируемость вариационного вывода

Долгое время Байесовские методы использовали для данных малого размера

- Слишком медленно для больших данных
- Все равно не очень выгодно

Все изменилось, когда Байесовские методы начали использовать в глубоком обучении, начали делать смешанные модели: нейронные сети вместе с вероятностной моделью.

Почему $p(x)$

Давайте приближать данный некоторым законом распределения $p(x)$

Почему $p(x)$

- Создание новых данных

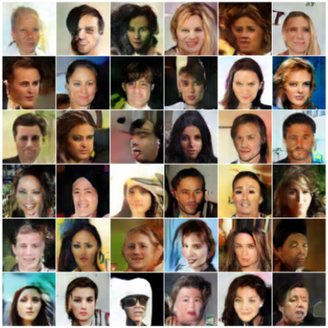


Рис.: https://arxiv.org/pdf/1609.03126.pdf

- Создание новых данных

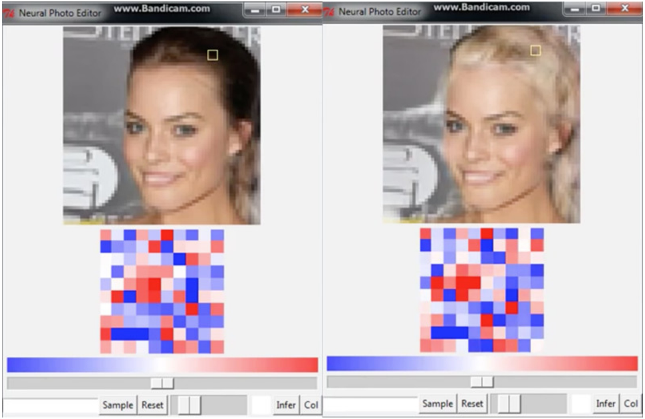


Рис.: См. Petapixal

Почему $p(x)$

- Создание новых данных
- Определение аномалий и выбросов (антифрод)
- Работа с пропусками в данных
- Представление данных в хорошем виде (например как p(молекула) для создания лекарств)

Как моделировать $p(x)$

- $\log \hat{p}(x) = CNN(x)$ — невозможно

$$p(x) = \frac{\exp(CNN)}{Z}$$

- $\log \hat{p}(x) = CNN(x)$ —невозможно
- Правило цепи (формула произведения вероятностей)

$$p(x_1, \dots, x_d) = p(x_1)p(x_2 \mid x_1) \cdots p(x_d \mid x_1, \dots, x_{d-1})$$

Рис.: Oord, Aaron van den, Nal Kalchbrenner, and Koray Kavukcuoglu. "Pixel recurrent neural networks." (2016)

Как моделировать $p(x)$

- $\log \hat{p}(x) = CNN(x)$ — невозможно
- Правило цепи (формула произведения вероятностей)

$$p(x_1, \ldots, x_d) = p(x_1)p(x_2 \mid x_1) \cdots p(x_d \mid x_1, \ldots, x_{d-1})$$
$$p(x_k \mid x_1, \ldots x_{k-1}) = RNN(x_1, \ldots, x_{k-1})$$

Рис.: Oord, Aaron van den, Nal Kalchbrenner, and Koray Kavukcuoglu. "Pixel recurrent neural networks." (2016)

Как моделировать $p(x)$

- $\log \hat{p}(x) = CNN(x)$ —невозможно
- Правило цепи (формула произведения вероятностей) —Хорошо, но долго генерировать
- $p(x_1, \ldots, x_d) = p(x_1) \cdots p(x_d)$

Как моделировать $p(x)$

- $\log \hat{p}(x) = CNN(x)$ —невозможно
- Правило цепи (формула произведения вероятностей) —Хорошо, но долго генерировать
- $p(x_1, \ldots, x_d) = p(x_1) \cdots p(x_d)$

Как моделировать $p(x)$

- $\log \hat{p}(x) = CNN(x)$ — невозможно
- Правило цепи (формула произведения вероятностей) —Хорошо, но долго генерировать
- $p(x_1, \dots, x_d) = p(x_1) \cdots p(x_d)$ — слишком ограничительно
- Смесь из нескольких гауссиан (GMM)

Как моделировать $p(x)$

- $\log \hat{p}(x) = CNN(x)$ — невозможно
- Правило цепи (формула произведения вероятностей) —Хорошо, но долго генерировать
- $p(x_1, \ldots, x_d) = p(x_1) \cdots p(x_d)$ — слишком ограничительно
- Смесь из нескольких гауссиан (GMM) — остается ограничительным
- Смесь бесконечного кол-ва Гауссиан (PPCA):

$$p(x) = \int p(x \mid t) p(t) \, dt$$

Непрерывная смесь Гауссиан

$$p(x) = \int p(x \mid t) p(t)\, dt$$

$$p(t) = \mathcal{N}(0, I)$$

Непрерывная смесь Гауссиан

$$p(x) = \int p(x \mid t) p(t)\, dt$$
$$p(t) = \mathcal{N}(0, I)$$
$$p(x \mid t) = \mathcal{N}(\mu(t), \Sigma(t))$$

- Если $\mu(t) = Wt + b$, $\Sigma(t) = \Sigma_0$ — это PPCA
- Если x — изображение, пусть

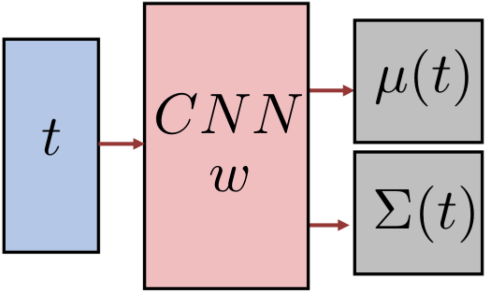
$$\mu(t) = CNN_1(t), \quad \Sigma(t) = CNN_2(t)$$

Непрерывная смесь Гауссиан

$$p(x \mid w) = \int p(x \mid t, w) p(t) \, dt$$

$$p(t) = \mathcal{N}(0, I)$$

$$p(x \mid t, w) = \mathcal{N}(\mu(t, w), \Sigma(t, w))$$

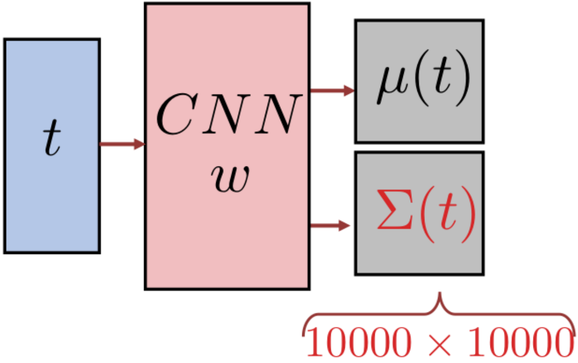


Непрерывная смесь Гауссиан

$$p(x \mid w) = \int p(x \mid t, w) p(t)\, dt$$

$$p(t) = \mathcal{N}(0, I)$$

$$p(x \mid t, w) = \mathcal{N}(\mu(t, w), \Sigma(t, w))$$

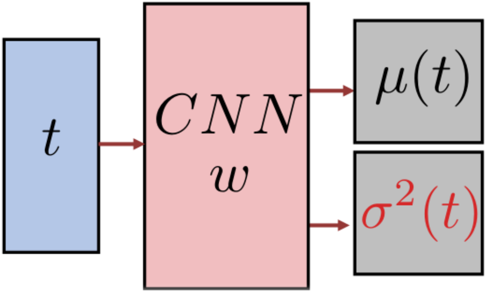


Непрерывная смесь Гауссиан

$$p(x \mid w) = \int p(x \mid t, w) p(t)\, dt$$

$$p(t) = \mathcal{N}(0, I)$$

$$p(x \mid t, w) = \mathcal{N}(\mu(t, w), diag(\sigma^2(t, w)))$$



Масштабирование EM

$$w \in \arg\max_w p(X \mid w), \quad p(X \mid w) = \int p(X \mid T, w) p(w)\, dt$$

Имеются латентные переменные $\Rightarrow$ используем EM!

$$\log p(X \mid w) \geq \mathcal{L}(w, q)$$
$$\max_{w,q} \mathcal{L}(w, q)$$

Масштабирование EM

$$w \in \arg\max_w p(X \mid w), \quad p(X \mid w) = \int p(X \mid T, w) p(w)\, dt$$

Имеются латентные переменные $\Rightarrow$ используем EM!

$$\log p(X \mid w) \geq \mathcal{L}(w, q)$$

$$\max_{w,q} \mathcal{L}(w, q)$$

Необходимо вычислить $p(T \mid X, w)$

Масштабирование EM

$$w \in \arg\max_w p(X \mid w), \quad p(X \mid w) = \int p(X \mid T, w) p(w)\, dt$$

Имеются латентные переменные $\Rightarrow$ используем EM!

MCMC? Можно сделать лучше

$$\mathrm{M}_q \log p(X, T \mid w) \approx \frac{1}{N} \sum_{k=1}^{N} \log(X, T_k \mid w)$$

$$T_k \sim q(T)$$

Масштабирование EM

$$w \in \arg\max_w p(X \mid w), \quad p(X \mid w) = \int p(X \mid T, w) p(w)\, dt$$

Имеются латентные переменные $\Rightarrow$ используем EM!

MCMC? Тогда вариационный EM!

$$\log p(X \mid w) \geq \mathcal{L}(w, q)$$
$$\max_{w,q} \mathcal{L}(w, q)$$

где $q_i(t_i) = \widetilde{q}(t_{i1}) \cdots \widetilde{q}(t_{im})$

Масштабирование вариационного EM

$$\max_{w,q} \mathcal{L}(w, q_1, \ldots, q_N)$$

$$\text{где } q_i(t_i) = \widetilde{q}(t_{i1}) \cdots \widetilde{q}(t_{im})$$

Масштабирование вариационного EM

$$\max_{\substack{w,m_1,\ldots,m_N \\ s_1,\ldots,s_N}} \mathcal{L}(w, q_1, \ldots, q_N)$$

$$\text{где } q_i(t_i) = \mathcal{N}(t_i \mid m_i, diag(s_i^2))$$

- Но в этом случае $\sim$ 100 параметров для каждого объекта
- Не ясно, что такое m, s для тестовых объектов

Масштабирование вариационного EM

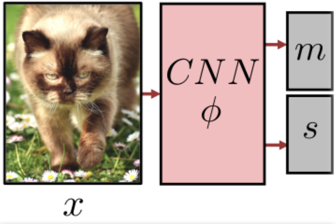
$$\max_{w,\phi} \mathcal{L}(w, q_1, \ldots, q_N)$$

где $q_i(t_i) = \mathcal{N}(t_i \mid m(x_i, \phi), diag(s^2(x_i, \phi)))$

Масштабирование вариационного EM

$$\max_{w,\phi} \mathcal{L}(w, q_1, \dots, q_N)$$

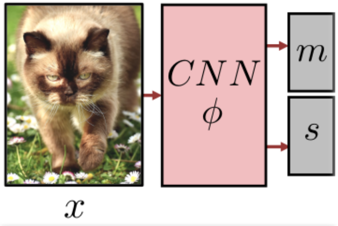
где $q_i(t_i) = \mathcal{N}(t_i \mid m(x_i, \phi), diag(s^2(x_i, \phi)))$



Масштабирование вариационного EM

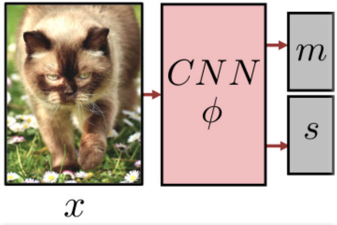
$$\max_{w,\phi} \sum_i \mathsf{M}_{q_i} \log \frac{p(x_i \mid t_i, w) p(t_i)}{q_i(t_i)}$$

где $q_i(t_i) = \mathcal{N}(t_i \mid m(x_i, \phi), diag(s^2(x_i, \phi)))$



$$\max_{w,\phi} \sum_i \mathsf{M}_{q_i} \log \frac{p(x_i \mid t_i, w) p(t_i)}{q_i(t_i)}$$

где $q_i(t_i) = \mathcal{N}(t_i \mid m(x_i, \phi), diag(s^2(x_i, \phi)))$

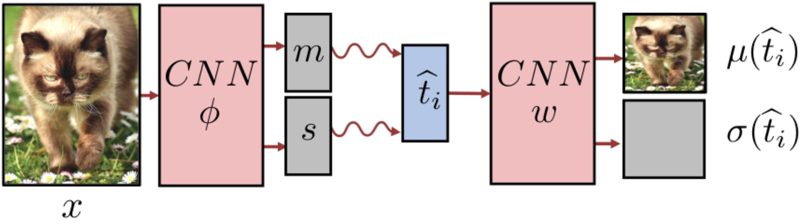


$$\hat{t}_i \sim \mathcal{N}(m(x_i, \phi), diag(s^2(x_i, \phi)))$$

Масштабирование вариационного EM

$$\max_{w,\phi} \sum_i \mathsf{M}_{q_i} \log \frac{p(x_i \mid t_i, w)p(t_i)}{q_i(t_i)}$$

где $q_i(t_i) = \mathcal{N}(t_i \mid m(x_i, \phi), diag(s^2(x_i, \phi)))$

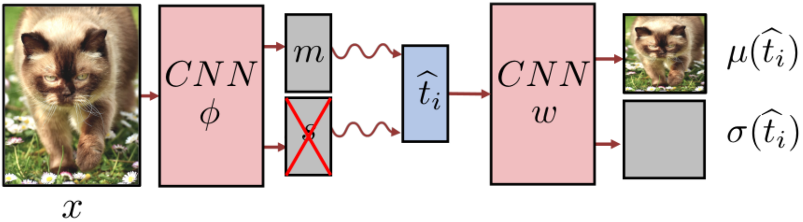


$$\hat{t}_i \sim \mathcal{N}(m(x_i, \phi), diag(s^2(x_i, \phi)))$$

Масштабирование вариационного EM

$$\max_{w,\phi} \sum_i \mathrm{M}_{q_i} \log \frac{p(x_i \mid t_i, w) p(t_i)}{q_i(t_i)}$$

где $q_i(t_i) = \mathcal{N}(t_i \mid m(x_i, \phi), diag(s^2(x_i, \phi)))$

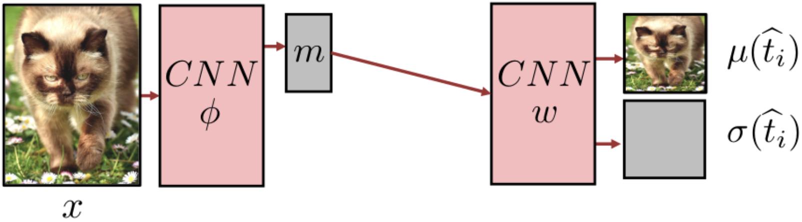


Если $s(x) = 0$, тогда $\hat{t}_i = m(x_i, \phi)$: обычный автокодировщик (autoencoder)

Масштабирование вариационного EM

$$\max_{w,\phi} \sum_i \mathrm{M}_{q_i} \log \frac{p(x_i \mid t_i, w)p(t_i)}{q_i(t_i)}$$

где $q_i(t_i) = \mathcal{N}(t_i \mid m(x_i, \phi), diag(s^2(x_i, \phi)))$



Если $s(x) = 0$, тогда $\hat{t}_i = m(x_i, \phi)$: обычный автокодировщик

Масштабирование вариационного EM

$$\max_{w,\phi} \sum_i \mathsf{M}_{q_i} \log \frac{p(x_i \mid t_i, w) p(t_i)}{q_i(t_i)} =$$

$$= \sum_i \mathsf{M}_{q_i} \log p(x_i \mid t_i, w) + \underbrace{\mathsf{M}_{q_i} \log \frac{p(t_i)}{q_i(t_i)}}_{-D_{KL}(q_i(t_i) \,\|\, p(t_i))}$$

Масштабирование вариационного EM

$$\max_{w,\phi} \sum_i \mathrm{M}_{q_i} \log \frac{p(x_i \mid t_i, w) p(t_i)}{q_i(t_i)} =$$

$$= \sum_i \mathrm{M}_{q_i} \underbrace{\log p(x_i \mid t_i, w)}_{-\|x_i - \mu(t_i)\|^2 + c} - D_{KL}(q_i(t_i) \parallel p(t_i))$$

$\uparrow$

Если $\overbrace{\sigma(x_i) = 1}$ для простоты

Масштабирование вариационного EM

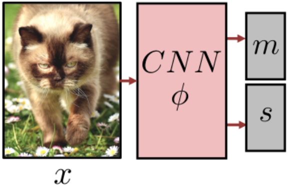
$$\max_{w,\phi} \sum_i \mathsf{M}_{q_i} \log \frac{p(x_i \mid t_i, w) p(t_i)}{q_i(t_i)} =$$

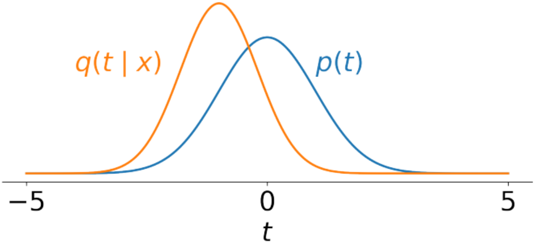
$$= \sum_i \mathsf{M}_{q_i} \underbrace{\log p(x_i \mid t_i, w)}_{-\|x_i - \mu(t_i)\|^2 + c} - D_{KL}(q_i(t_i) \parallel p(t_i))$$

где

- $\mathsf{M}_{q_i}(\cdot)$ — Аппроксимация апостериорного распределения $q(t_i) \approx p(t_i \mid x_i, w_i)$
- $-\|x_i - \mu(t_i)\|^2$ — Ошибка восстановления
- $-D_{KL}(q_i(t_i) \parallel p(t_i))$ — Регуляризация

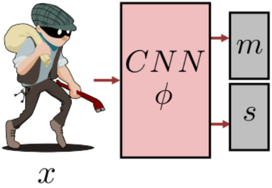
Обнаружение выбросов

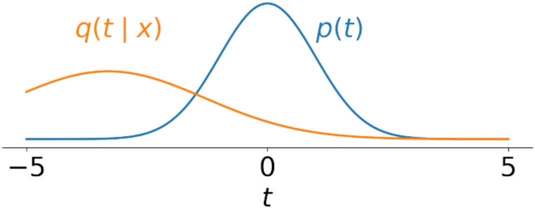




$$D_{KL}(q(t \mid s) \parallel p(t)) \approx 0.54$$

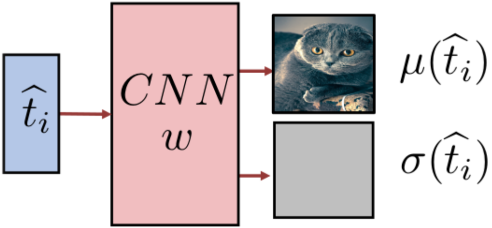
Обнаружение выбросов





$$D_{KL}(q(t \mid s) \parallel p(t)) \approx \color{red}{6.25}$$

$$p(x \mid w) = \int p(x \mid t, w) p(t)\, dt$$



$\hat{t}_i \sim \mathcal{N}(0, I)$

Вычисление градиента

$$\max_{w,\phi} \sum_i \mathrm{M}_{q_i} \log p(x_i \mid t_i, w) - D_{KL}(q_i(t_i) \parallel p(t_i))$$

Вычисление градиента

$$\max_{w,\phi} \sum_i \mathrm{M}_{q_i} \log p(x_i \mid t_i, w) - \underbrace{D_{KL}(q_i(t_i) \parallel p(t_i))}_{\text{Есть аналитическое в-ие}}$$

$$D_{KL}(q_i(t_i) \parallel p(t_i)) =$$
$$= \sum_i \left(-\log \sigma_j(t_i) + \frac{\sigma_j^2(t_i) + \mu_j^2(t_i) - 1}{2} \right)$$

Вычисление градиента

$$f(w, \phi) = \sum_i \mathrm{M}_{q_i} \log p(x_i \mid t_i, w)$$

Вычисление градиента

$$f(w, \phi) = \sum_i \mathrm{M}_{q_i} \log p(x_i \mid t_i, w)$$

$$q_i(t_i) = q(t_i \mid x_i, \phi_i) = \mathcal{N}(t_i \mid m_i, diag(s_i^2))$$

Вычисление градиента

$$f(w,\phi) = \sum_i \mathrm{M}_{q_i(t_i|x_i,\phi)} \log p(x_i \mid t_i, w)$$

$$q_i(t_i) = q(t_i \mid x_i, \phi_i) = \mathcal{N}(t_i \mid m_i, diag(s_i^2))$$

Вычисление градиента

$$\nabla_w f(w, \phi) = \nabla_w \sum_i \mathsf{M}_{q_i(t_i|x_i,\phi)} \log p(x_i \mid t_i, w)$$

Вычисление градиента

$$\nabla_w f(w, \phi) = \sum_i \int \nabla_w q_i(t_i \mid x_i, \phi) \log p(x_i \mid t_i, w) \, dt_i$$

Вычисление градиента

$$\nabla_w f(w, \phi) = \sum_i \int q_i(t_i \mid x_i, \phi) \nabla_w \log p(x_i \mid t_i, w)\, dt_i =$$
$$= \mathsf{M}_{q_i(t_i \mid x_i, \phi)} \nabla_w \log p(x_i \mid t_i, w) \approx$$
$$\approx \sum_i \nabla_w \log p(x_i \mid \hat{t}_i, w)$$

$$\hat{t}_i \sim q(t_i \mid x_i, \phi_i)$$

Вычисление градиента

$$\nabla_w f(w, \phi) \approx \sum_i \underbrace{\nabla_w \log p(x_i \mid \hat{t}_i, w)}_{\text{Градиент стандартной NN}}$$

$$\hat{t}_i \sim q(t_i \mid x_i, \phi_i)$$

Вычисление градиента

$$\nabla_w f(w, \phi) \approx \sum_i \nabla_w \log p(x_i \mid \hat{t}_i, w) \approx$$

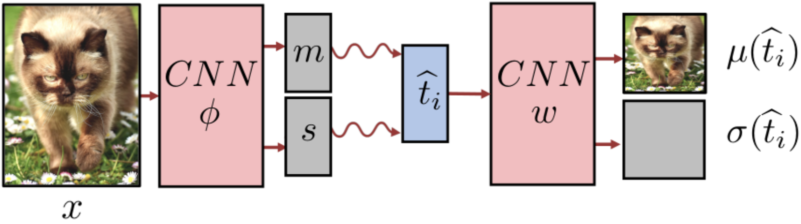
$$\approx \frac{N}{n} \sum_{k=1}^{n} \underbrace{\nabla_w \log p(x_{i_k} \mid \hat{t}_{i_k}, w)}_{\text{Стох. градиент стандартной NN}}$$

$$\hat{t}_i \sim q(t_i \mid x_i, \phi_i)$$

Вычисление градиента

$$\nabla_w f(w,\phi) \approx \sum_i \nabla_w \log p(x_i \mid \hat{t}_i, w) \approx$$

$$\approx \frac{N}{n} \sum_{k=1}^{n} \underbrace{\nabla_w \log p(x_{i_k} \mid \hat{t}_{i_k}, w)}_{\text{Стох. градиент стандартной NN}}$$



$$\hat{t}_i \sim \mathcal{N}(m(x_i,\phi), diag(s^2(x_i,\phi)))$$

Вычисление градиента

$$\nabla_{\phi} f(w, \phi) = \nabla_{\phi} \sum_i \mathsf{M}_{q_i(t_i \mid x_i, \phi)} \log p(x_i \mid t_i, w)$$

Вычисление градиента

$$\nabla_\phi f(w,\phi) = \color{red}{\nabla_\phi} \color{black} \sum_i \int q_i(t_i \mid x_i, \phi) \log p(x_i \mid t_i, w)\, dt_i$$

Вычисление градиента

$$\nabla_\phi f(w, \phi) = \sum_i \int \nabla_\phi q_i(t_i \mid x_i, \phi) \log p(x_i \mid t_i, w)\, dt_i$$

Вычисление градиента

$$\nabla_\phi f(w,\phi) = \sum_i \int \nabla_\phi q_i(t_i \mid x_i,\phi) \log p(x_i \mid t_i, w)\, dt_i =$$

$$= \sum_i \int \frac{q(t_i \mid x_i,\phi)}{q(t_i \mid x_i,\phi)} \nabla_\phi q_i(t_i \mid x_i,\phi) \log p(x_i \mid t_i, w)\, dt_i$$

$$\boxed{\nabla \log g(\phi) = \frac{\nabla g(\phi)}{g(\phi)}}$$

Вычисление градиента

$$\nabla_\phi f(w,\phi) = \sum_i \int \nabla_\phi q_i(t_i \mid x_i, \phi) \log p(x_i \mid t_i, w)\, dt_i =$$

$$= \sum_i \int \frac{q(t_i \mid x_i, \phi)}{q(t_i \mid x_i, \phi)} \nabla_\phi q_i(t_i \mid x_i, \phi) \log p(x_i \mid t_i, w)\, dt_i =$$

$$= \sum_i \int q(t_i \mid x_i, \phi) \nabla_\phi \log q_i(t_i \mid x_i, \phi) \log p(x_i \mid t_i, w)\, dt_i$$

$$\boxed{\nabla \log g(\phi) = \frac{\nabla g(\phi)}{g(\phi)}}$$

Вычисление градиента

$$\nabla_\phi f(w,\phi) = \sum_i \int \nabla_\phi q_i(t_i \mid x_i, \phi) \log p(x_i \mid t_i, w)\, dt_i =$$

$$= \sum_i \int \frac{q(t_i \mid x_i, \phi)}{q(t_i \mid x_i, \phi)} \nabla_\phi q_i(t_i \mid x_i, \phi) \log p(x_i \mid t_i, w)\, dt_i =$$

$$= \sum_i \int q(t_i \mid x_i, \phi) \nabla_\phi \log q_i(t_i \mid x_i, \phi) \log p(x_i \mid t_i, w)\, dt_i =$$

$$= \sum_i \int \mathrm{M}_{q(t_i|x_i,\phi)} \nabla_\phi \log q_i(t_i \mid x_i, \phi) \log p(x_i \mid t_i, w)\, dt_i$$

Вычисление градиента

$$\nabla_\phi f(w,\phi) = \sum_i \int \nabla_\phi q_i(t_i \mid x_i, \phi) \log p(x_i \mid t_i, w)\, dt_i =$$

$$= \sum_i \int \frac{q(t_i \mid x_i, \phi)}{q(t_i \mid x_i, \phi)} \nabla_\phi q_i(t_i \mid x_i, \phi) \log p(x_i \mid t_i, w)\, dt_i =$$

$$= \sum_i \int q(t_i \mid x_i, \phi) \nabla_\phi \log q_i(t_i \mid x_i, \phi) \log p(x_i \mid t_i, w)\, dt_i =$$

$$= \sum_i \int \mathrm{M}_{q(t_i \mid x_i, \phi)} \nabla_\phi \log q_i(t_i \mid x_i, \phi) \underbrace{\log p(x_i \mid t_i, w)}_{-inf}\, dt_i$$

Вычисление градиента

$$\nabla_\phi f(w,\phi) = \nabla_\phi \sum_i \mathrm{M}_{q_i(t_i|x_i,\phi)} \log p(x_i \mid t_i, w)$$

$$t_i \sim q(t_i \mid x_i, \phi) = \mathcal{N}(t_i \mid m_i, diag(s_i^2))$$
$$t_i = \varepsilon \cdot s_i + m_i = g(\varepsilon_i, x_i, \phi)$$
$$\varepsilon_i \sim p(\varepsilon_i) = \mathcal{N}(\varepsilon_i \mid O, I)$$

Вычисление градиента

$$\nabla_\phi f(w,\phi) = \nabla_\phi \sum_i \mathrm{M}_{q_i(t_i|x_i,\phi)} \log p(x_i \mid t_i, w) =$$

$$= \sum_i \nabla_\phi \mathrm{M}_{p(\varepsilon_i)} \log p(x_i \mid g(\varepsilon_i, x_i, \phi), w)$$

$$t_i \sim q(t_i \mid x_i, \phi) = \mathcal{N}(t_i \mid m_i, diag(s_i^2))$$

$$t_i = \varepsilon \cdot s_i + m_i = g(\varepsilon_i, x_i, \phi)$$

$$\varepsilon_i \sim p(\varepsilon_i) = \mathcal{N}(\varepsilon_i \mid O, I)$$

Вычисление градиента

$$\nabla_\phi f(w,\phi) = \nabla_\phi \sum_i \mathrm{M}_{q_i(t_i|x_i,\phi)} \log p(x_i \mid t_i, w) =$$

$$= \sum_i \nabla_\phi \mathrm{M}_{p(\varepsilon_i)} \log p(x_i \mid g(\varepsilon_i, x_i, \phi), w) =$$

$$= \sum_i \int p(\varepsilon_i) \nabla_\phi \log p(x_i \mid g(\varepsilon_i, x_i, \phi), w)$$

$$t_i \sim q(t_i \mid x_i, \phi) = \mathcal{N}(t_i \mid m_i, diag(s_i^2))$$

$$t_i = \varepsilon \cdot s_i + m_i = g(\varepsilon_i, x_i, \phi)$$

$$\varepsilon_i \sim p(\varepsilon_i) = \mathcal{N}(\varepsilon_i \mid 0, I)$$

Вычисление градиента

$$\nabla_\phi f(w,\phi) = \nabla_\phi \sum_i \mathsf{M}_{q_i(t_i|x_i,\phi)} \log p(x_i \mid t_i, w) =$$

$$= \sum_i \int p(\varepsilon_i) \nabla_\phi \log p(x_i \mid g(\varepsilon_i, x_i, \phi), w) =$$

$$= \sum_i \mathsf{M}_{p(\varepsilon_i)} \nabla_\phi \log p(x_i \mid g(\varepsilon_i, x_i, \phi), w)$$

$$t_i \sim q(t_i \mid x_i, \phi) = \mathcal{N}(t_i \mid m_i, diag(s_i^2))$$

$$t_i = \varepsilon \cdot s_i + m_i = g(\varepsilon_i, x_i, \phi)$$

$$\varepsilon_i \sim p(\varepsilon_i) = \mathcal{N}(\varepsilon_i \mid 0, I)$$

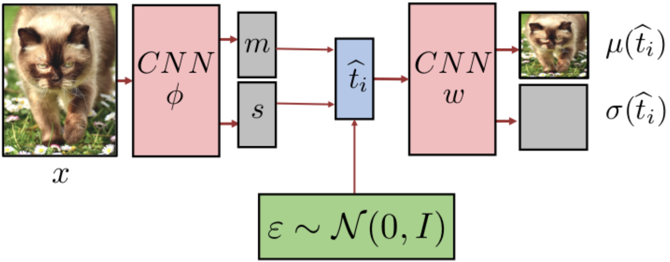
Вычисление градиента

$$\nabla_\phi f(w,\phi) = \sum_i \mathrm{M}_{p(\varepsilon_i)} \nabla_\phi \log p(x_i \mid g(\varepsilon_i, x_i, \phi), w)$$

$$t_i \sim q(t_i \mid x_i, \phi) = \mathcal{N}(t_i \mid m_i, diag(s_i^2))$$

$$t_i = \varepsilon \cdot s_i + m_i = g(\varepsilon_i, x_i, \phi)$$

$$\varepsilon_i \sim p(\varepsilon_i) = \mathcal{N}(\varepsilon_i \mid O, I)$$



Вариационный автокодировщик

- Бесконечная смесь Гауссиан
- Для обучения: EM + аппроксимация q гауссианами + стохастический вариационный вывод
- Как обычный автокодировщик (autoencoder), но только с шумом и KL регуляризацией
- Моделирует отличные картинки